

## LA-UR-21-27531

Approved for public release; distribution is unlimited.

Title: A Machine Learning aided hierarchical screening strategy for materials discovery

Author(s): Talapatra, Anjana Anu

Intended for: Slides to accompany tutorial (already approved) on nanoHUB.org

Issued: 2021-07-30

---

**Disclaimer:**

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

# A Machine Learning aided hierarchical screening strategy for materials discovery

Anjana Talapatra

Director's Postdoctoral Fellow  
Materials Science & Technology Division  
Los Alamos National Laboratory

21<sup>st</sup> July, 2021



# Discovery and Design of Novel wide band gap **Oxide Perovskites**

1  
H  
Hydrogen  
Nonmetal

3  
Li  
Lithium  
Alkali Metal

11  
Na  
Sodium  
Alkali Metal

19  
K  
Potassium  
Alkali Metal

55  
Cs  
Cesium  
Alkali Metal

87  
Fr  
Francium  
Alkali Metal

4  
Be  
Beryllium  
Alkaline Earth Metal

12  
Mg  
Magnesium  
Alkaline Earth Metal

20  
Ca  
Calcium  
Alkaline Earth Metal

56  
Ba  
Barium  
Alkaline Earth Metal

88  
Ra  
Radium  
Alkaline Earth Metal

21  
Sc  
Scandium  
Transition Metal

39  
Y  
Yttrium  
Transition Metal

57  
La  
Lanthanum  
Lanthanide

89  
Ac  
Actinium  
Actinide

22  
Ti  
Titanium  
Transition Metal

40  
Zr  
Zirconium  
Transition Metal

58  
Ce  
Cerium  
Lanthanide

90  
Th  
Thorium  
Actinide

23  
V  
Vanadium  
Transition Metal

41  
Nb  
Niobium  
Transition Metal

59  
Pr  
Praseodymium  
Lanthanide

91  
Pa  
Protactinium  
Actinide

24  
Cr  
Chromium  
Transition Metal

42  
Mo  
Molybdenum  
Transition Metal

60  
Nd  
Neodymium  
Lanthanide

92  
U  
Uranium  
Actinide

25  
Mn  
Manganese  
Transition Metal

43  
Tc  
Technetium  
Transition Metal

61  
Pm  
Promethium  
Lanthanide

93  
Np  
Neptunium  
Actinide

26  
Fe  
Iron  
Transition Metal

44  
Ru  
Ruthenium  
Transition Metal

62  
Sm  
Samarium  
Lanthanide

94  
Pu  
Plutonium  
Actinide

27  
Co  
Cobalt  
Transition Metal

45  
Rh  
Rhodium  
Transition Metal

63  
Eu  
Europium  
Lanthanide

95  
Am  
Americium  
Actinide

28  
Ni  
Nickel  
Transition Metal

46  
Pd  
Palladium  
Transition Metal

64  
Gd  
Gadolinium  
Lanthanide

96  
Cm  
Curium  
Actinide

29  
Cu  
Copper  
Transition Metal

47  
Ag  
Silver  
Transition Metal

65  
Tb  
Terbium  
Lanthanide

97  
Bk  
Berkelium  
Actinide

30  
Zn  
Zinc  
Transition Metal

48  
Cd  
Cadmium  
Transition Metal

66  
Dy  
Dysprosium  
Lanthanide

98  
Cf  
Californium  
Actinide

31  
Ga  
Gallium  
Post-Transition Metal

49  
In  
Indium  
Post-Transition Metal

67  
Ho  
Holmium  
Lanthanide

99  
Es  
Einsteinium  
Actinide

32  
Ge  
Germanium  
Metalloid

50  
Sn  
Tin  
Post-Transition Metal

68  
Er  
Erbium  
Lanthanide

100  
Fm  
Fermium  
Actinide

33  
As  
Arsenic  
Metalloid

51  
Sb  
Antimony  
Metalloid

69  
Tm  
Thulium  
Lanthanide

101  
Md  
Mendelevium  
Actinide

34  
Se  
Selenium  
Nonmetal

52  
Te  
Tellurium  
Metalloid

70  
Yb  
Ytterbium  
Lanthanide

102  
No  
Nobelium  
Actinide

35  
Br  
Bromine  
Halogen

53  
I  
Iodine  
Halogen

71  
Lu  
Lutetium  
Lanthanide

103  
Lr  
Lawrencium  
Actinide

36  
Kr  
Krypton  
Noble Gas

54  
Xe  
Xenon  
Noble Gas

72  
Be  
Beryllium  
Alkaline Earth Metal

104  
Rf  
Rutherfordium  
Transition Metal

37  
Rb  
Rubidium  
Alkali Metal

55  
Cs  
Cesium  
Alkali Metal

73  
La  
Lanthanum  
Lanthanide

105  
Db  
Dubnium  
Transition Metal

38  
Sr  
Strontium  
Alkaline Earth Metal

56  
Ba  
Barium  
Alkaline Earth Metal

74  
Ce  
Cerium  
Lanthanide

106  
Sg  
Seaborgium  
Transition Metal

39  
Y  
Yttrium  
Transition Metal

57  
La  
Lanthanum  
Lanthanide

75  
Re  
Rhenium  
Transition Metal

107  
Bh  
Bohrium  
Transition Metal

40  
Zr  
Zirconium  
Transition Metal

58  
Ce  
Cerium  
Lanthanide

76  
Os  
Osmium  
Transition Metal

108  
Hs  
Hassium  
Transition Metal

41  
Nb  
Niobium  
Transition Metal

59  
Pr  
Praseodymium  
Lanthanide

77  
Ir  
Iridium  
Transition Metal

109  
Mt  
Meitnerium  
Transition Metal

42  
Mo  
Molybdenum  
Transition Metal

60  
Nd  
Neodymium  
Lanthanide

78  
Pt  
Platinum  
Transition Metal

110  
Ds  
Darmstadtium  
Transition Metal

43  
Tc  
Technetium  
Transition Metal

61  
Pm  
Promethium  
Lanthanide

79  
Au  
Gold  
Transition Metal

111  
Rg  
Roentgenium  
Transition Metal

44  
Ru  
Ruthenium  
Transition Metal

62  
Sm  
Samarium  
Lanthanide

80  
Hg  
Mercury  
Transition Metal

112  
Cn  
Copernicium  
Transition Metal

45  
Rh  
Rhodium  
Transition Metal

63  
Eu  
Europium  
Lanthanide

81  
Tl  
Thallium  
Post-Transition Metal

113  
Nh  
Nihonium  
Post-Transition Metal

46  
Pd  
Palladium  
Transition Metal

64  
Gd  
Gadolinium  
Lanthanide

82  
Pb  
Lead  
Post-Transition Metal

114  
Fl  
Flerovium  
Post-Transition Metal

47  
Ag  
Silver  
Transition Metal

65  
Tb  
Terbium  
Lanthanide

83  
Bi  
Bismuth  
Post-Transition Metal

115  
Mc  
Moscovium  
Post-Transition Metal

48  
Cd  
Cadmium  
Transition Metal

66  
Dy  
Dysprosium  
Lanthanide

84  
Po  
Polonium  
Metalloid

116  
Lv  
Livermorium  
Post-Transition Metal

49  
In  
Indium  
Post-Transition Metal

67  
Ho  
Holmium  
Lanthanide

85  
At  
Astatine  
Halogen

117  
Ts  
Tennessine  
Halogen

50  
Sn  
Tin  
Post-Transition Metal

68  
Er  
Erbium  
Lanthanide

86  
Rn  
Radon  
Noble Gas

118  
Og  
Oganesson  
Noble Gas

2  
He  
Helium  
Noble Gas

10  
Ne  
Neon  
Noble Gas

18  
Ar  
Argon  
Noble Gas

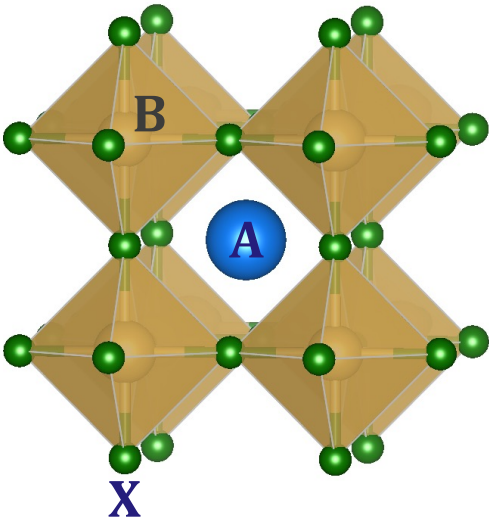
36  
Kr  
Krypton  
Noble Gas

54  
Xe  
Xenon  
Noble Gas

86  
Rn  
Radon  
Noble Gas

118  
Og  
Oganesson  
Noble Gas

Truly Vast Chemical Space  
(Millions of Combinatorial Possibilities)

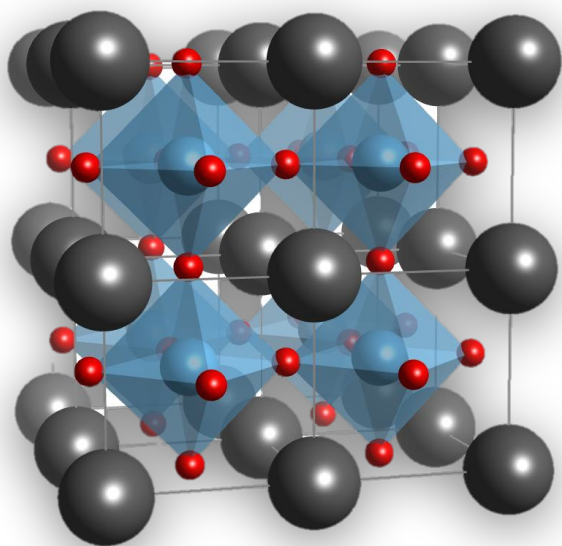




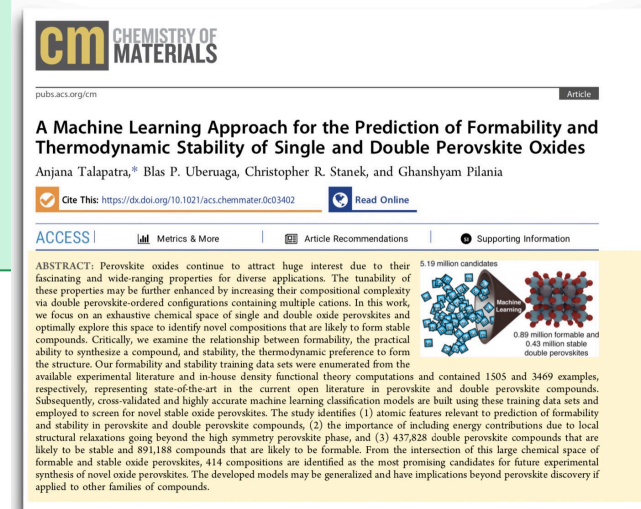
# Objective:

Down select from millions of potential compounds to a relatively small and tractable set of promising wide band gap oxide perovskites

## Oxide Perovskites



- The perovskite structure can accommodate 90% of the metallic ions in the periodic table
- Amenable to band gap tuning
- Exhibit fascinating electrical and magnetic properties:
  - piezoelectricity, optical properties,
  - high-temperature superconductivity,
  - ferroelectricity, magneto-strictive effects

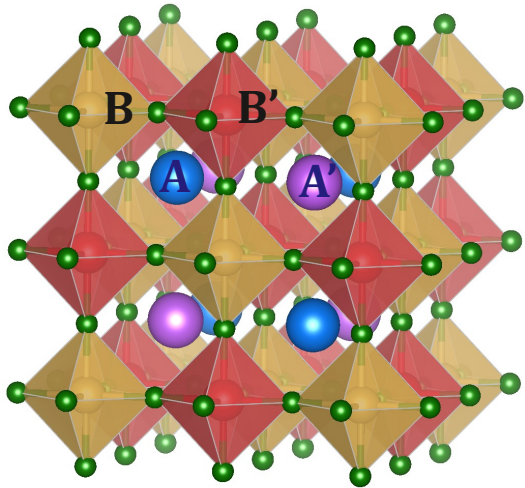


Talapatra, Anjana, et al. "A Machine Learning Approach for the Prediction of Formability and Thermodynamic Stability of Single and Double Perovskite Oxides." *Chemistry of Materials* (2021).

# A Strategy for Scintillator Discovery and Optimization

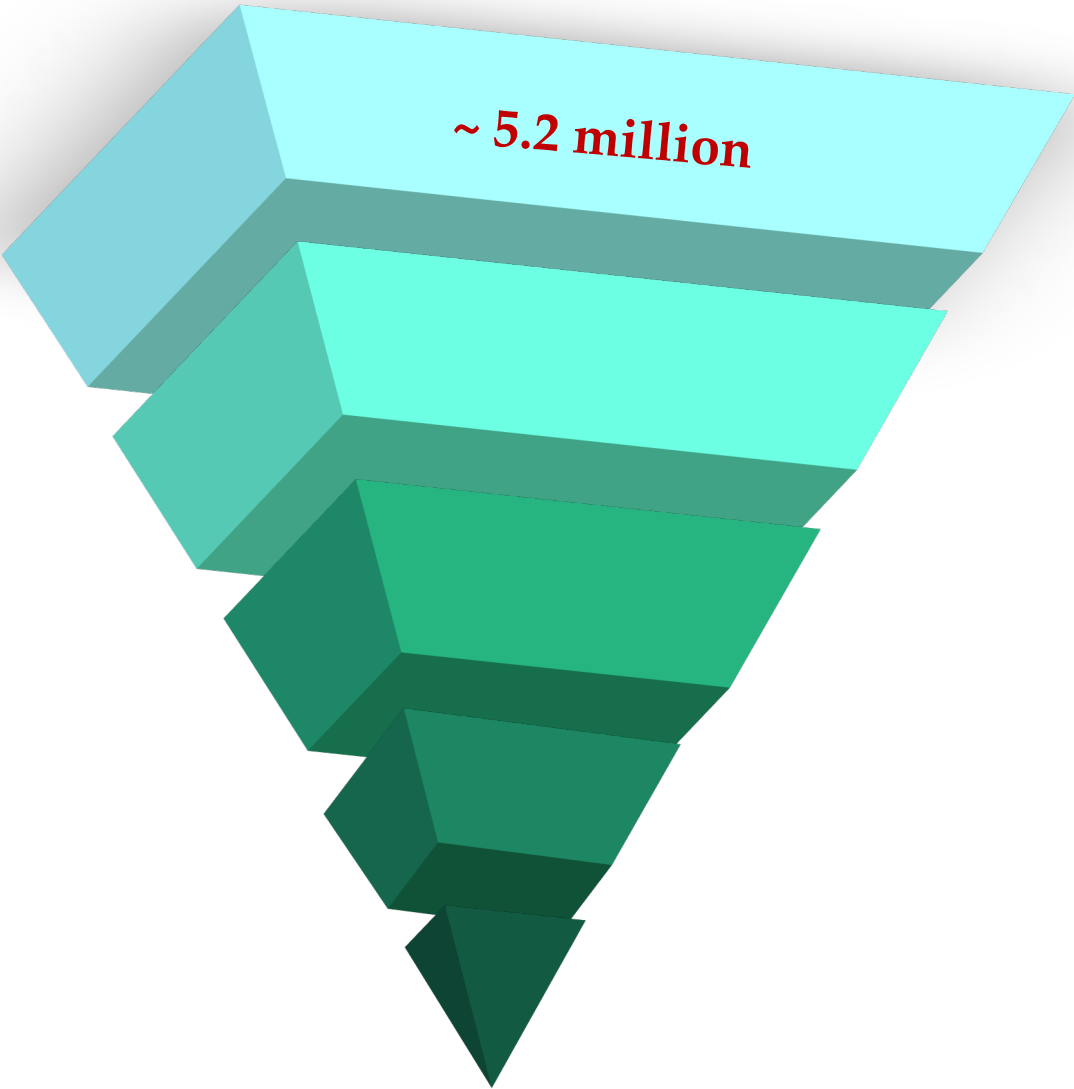
1 H																	2 He																		
3 Li	4 Be																	5 B	6 C	7 N	8 O	9 F	10 Ne												
11 Na	12 Mg																	13 Al	14 Si	15 P	16 S	17 Cl	18 Ar												
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	71 Lu	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn	87 Fr	88 Ra	103 Lr	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og
		57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb																				

68 elements



A	A'	B	B'
68	66	67	65

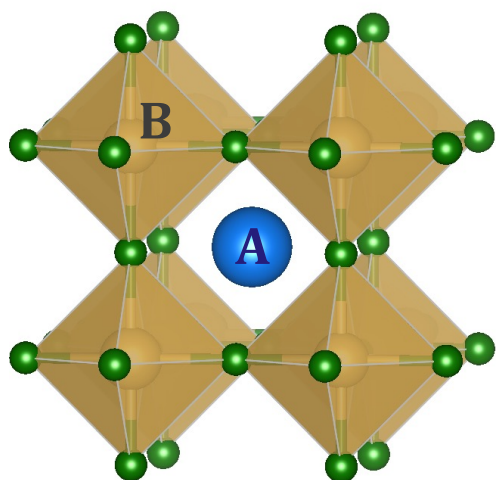
~ 5.2 million  
unique combinations



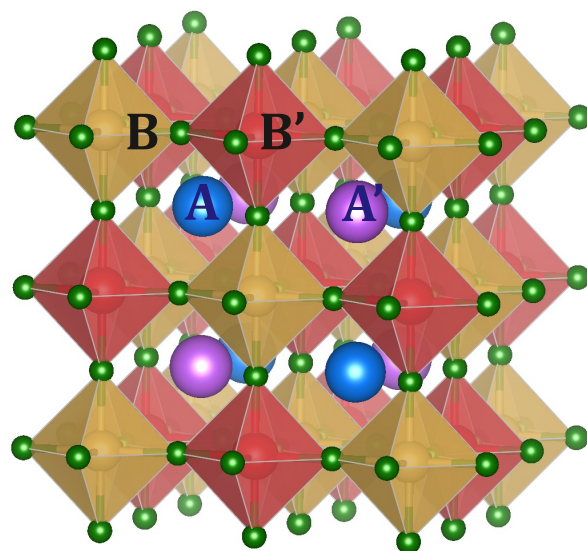
# Objective:

Down select from millions of potential compounds to a relatively small and tractable set of promising scintillators

Single perovskite ( $ABO_3$ )



Double perovskites



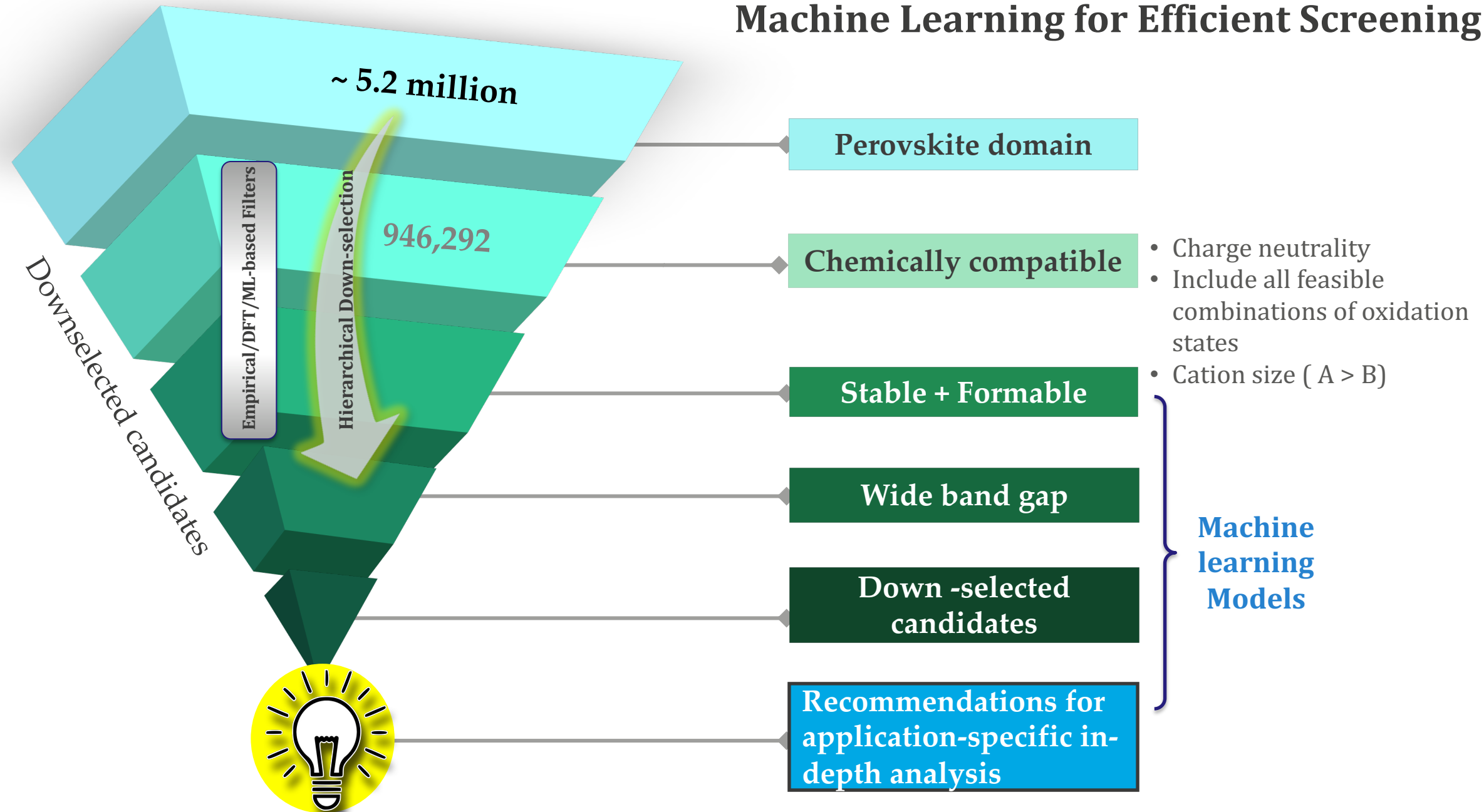
- $AA'B_2O_6$
- $A_2BB'O_6$
- $AA'BB'O_6$

- Assumptions:
  - 50-50 compositions
  - 2 elements per cation site
  - Cubic structures
  - Rocksalt ordering

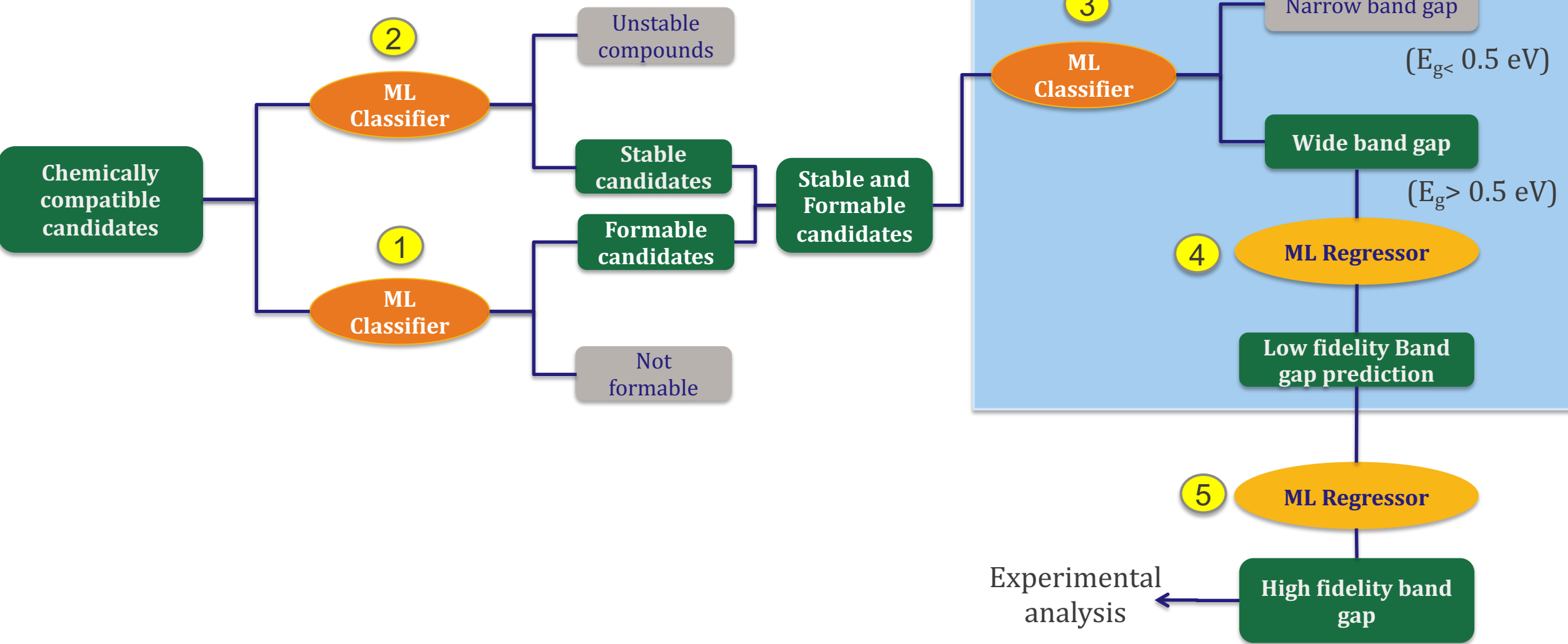
# Why Machine Learning?

- To screen millions of compounds
- Unearth relationships between electronic structure, chemistry, thermodynamic stability, formability and band gap
- We know how to:
  - Calculate thermodynamic stability
  - Calculate approximate band gap
- Very complicated to:
  - Estimate synthesizability

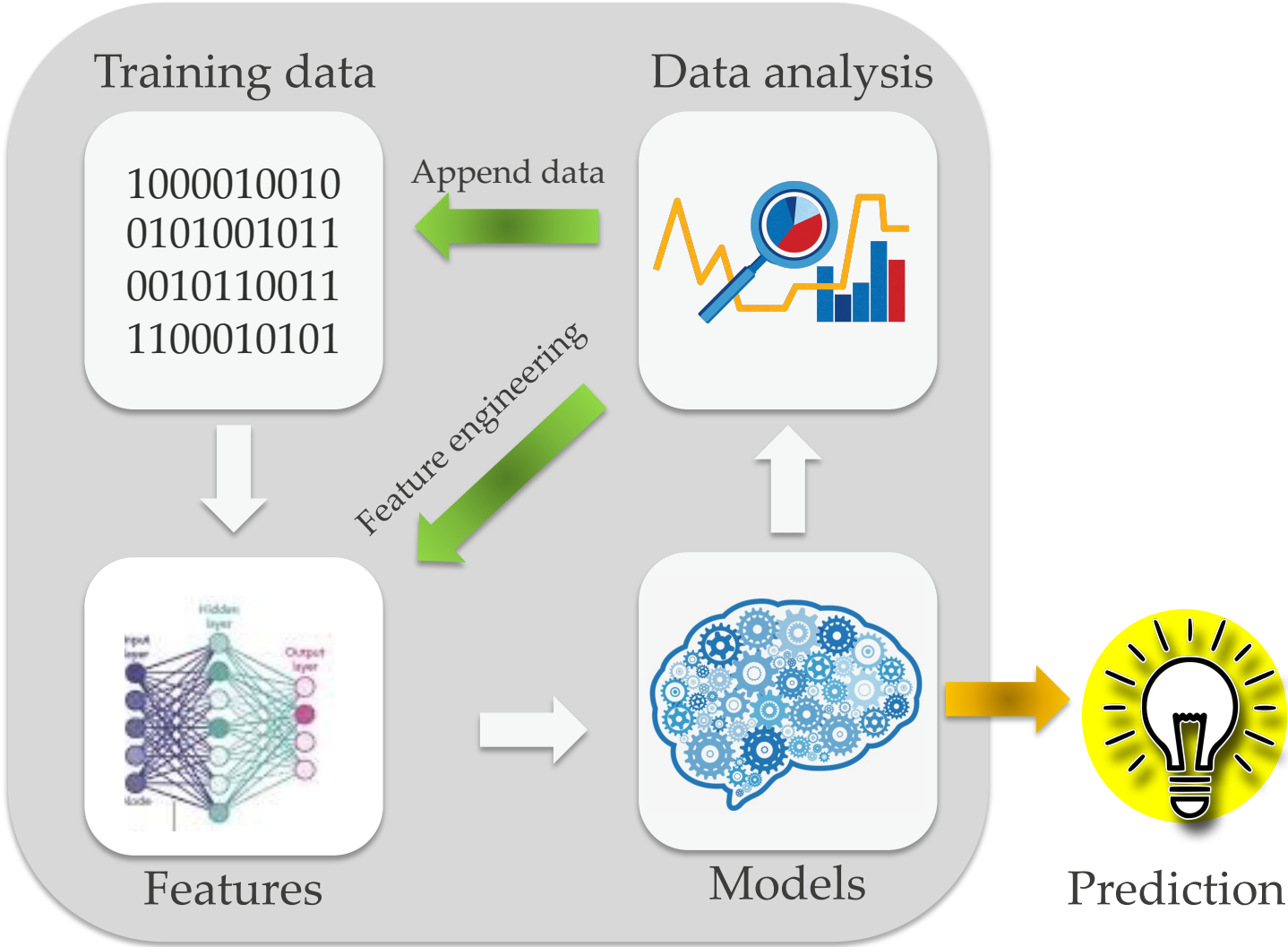
# Machine Learning for Efficient Screening



# Perovskite discovery using Machine Learning

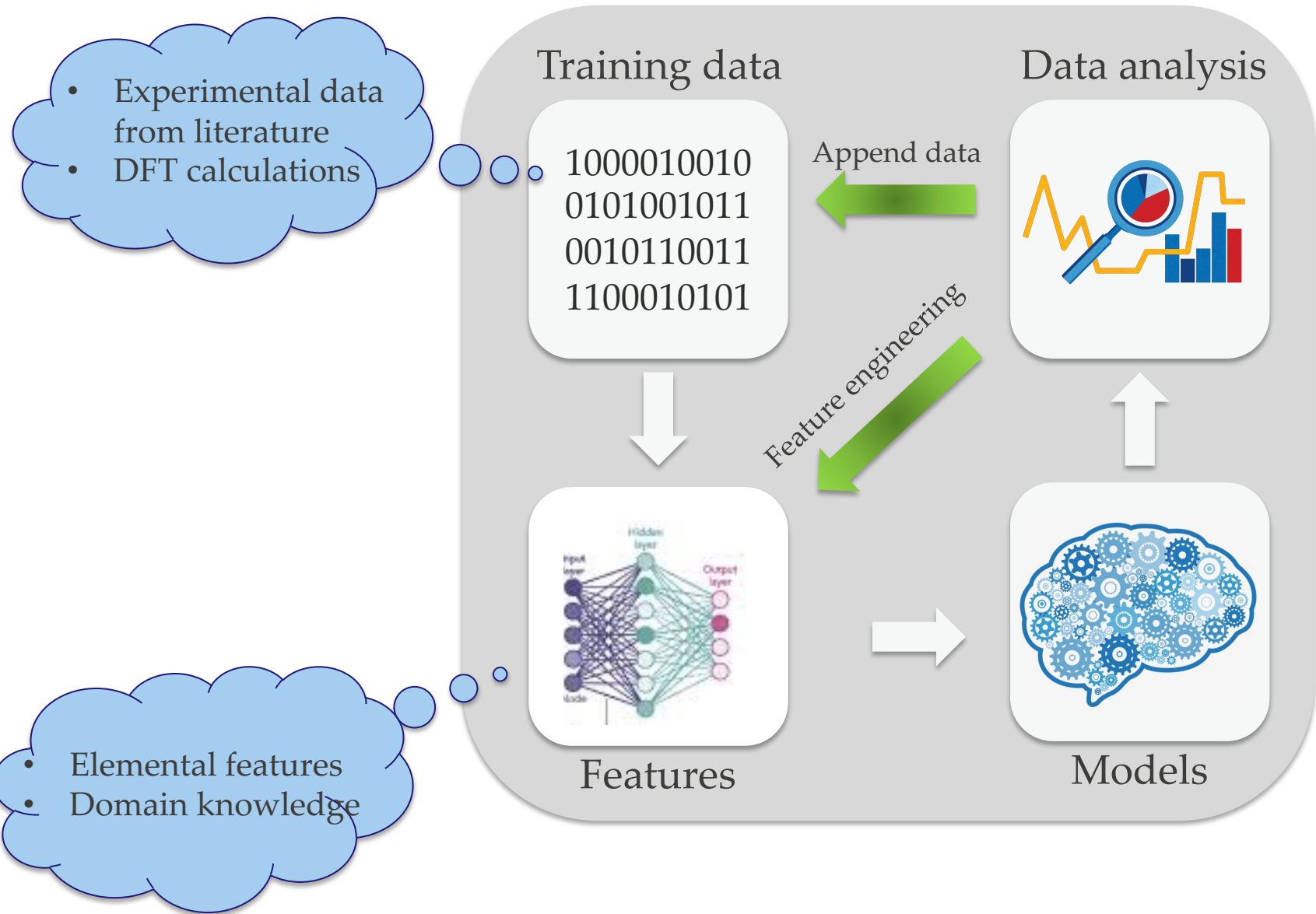


# Components of ML infrastructure





# Components of ML infrastructure

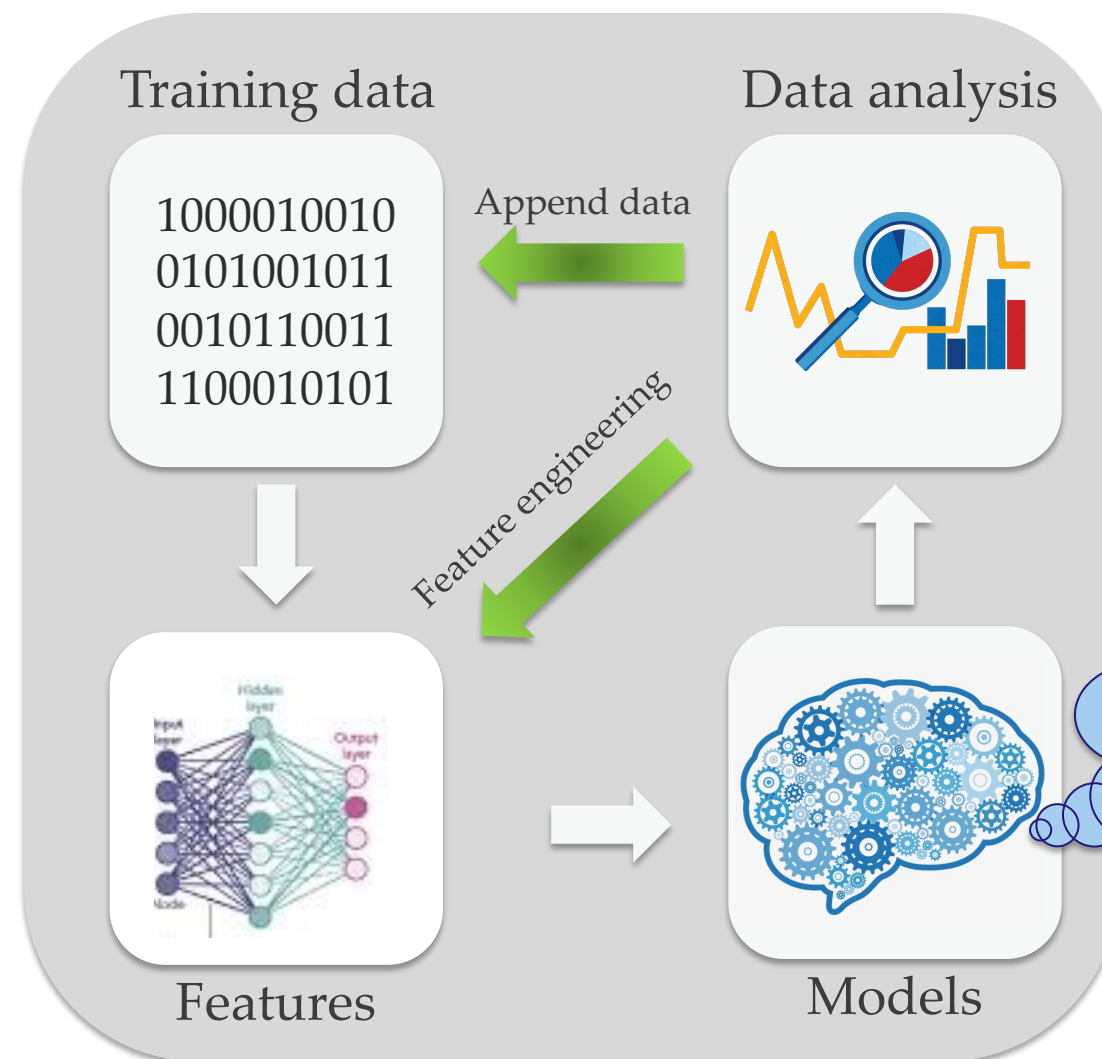


- Model is as good as your training data
- To increase applicability, ideally use features that are easy to populate
- Double check, triple check source of features and values of features.
- 90 % of bugs can be attributed to mistakes in populating training data
- Reproducibility and consistency



# Components of ML infrastructure

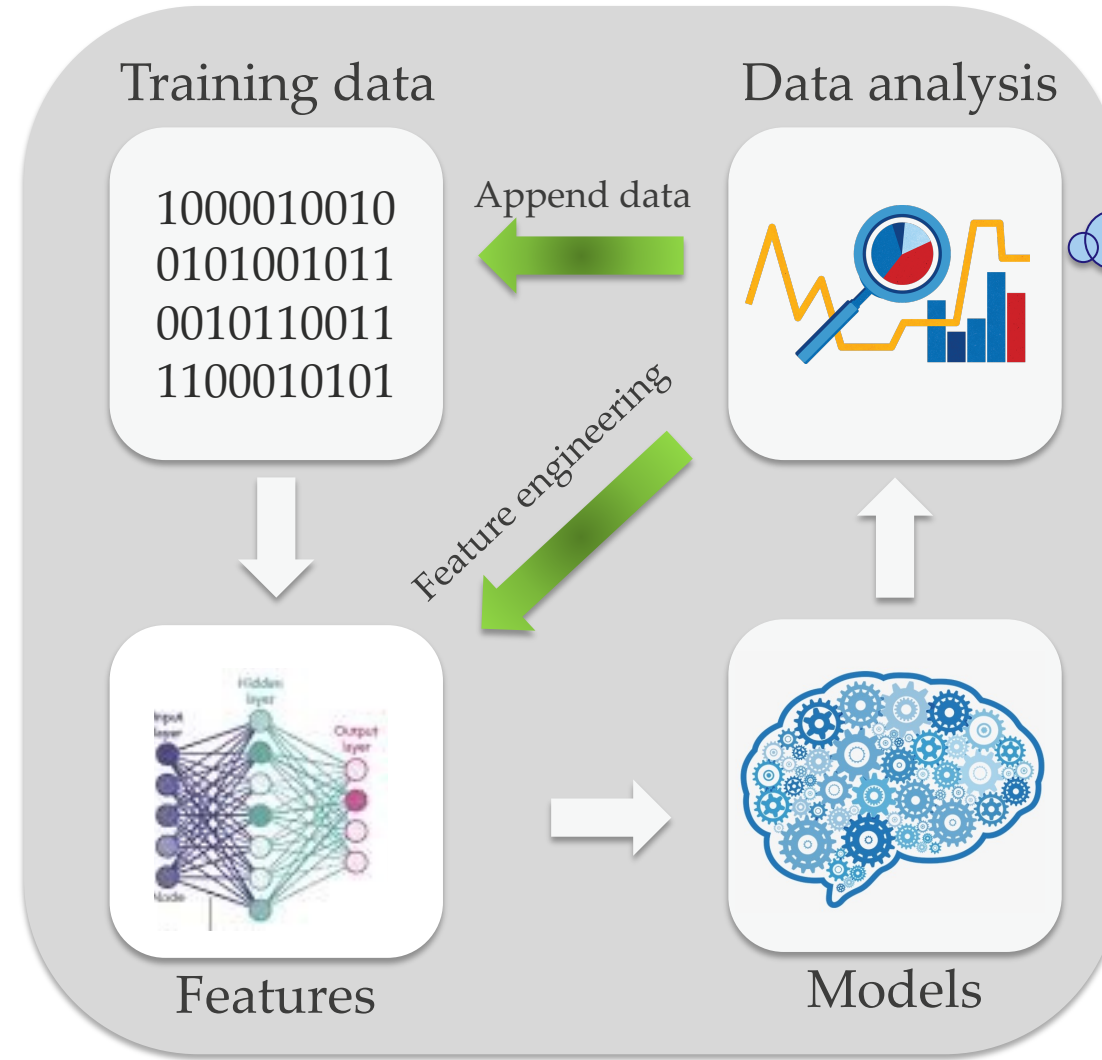
- Choose modeling technique wisely
- Baseline comparisons across different models
- Ensure that what you are producing is better than what is available, else use what is available



Random forest models:  
 - Classification (down-selection)  
 - Regression (band gap)

# Components of ML infrastructure

- If possible, increase training data adaptively keeping an eye on performance metrics
- Analyze data and results continuously to ensure it makes sense intuitively
- Avoid data leakage
- If something seems off, it probably is off.



Cross-validation  
Feature selection  
Testing  
Performance curves  
Partial dependence plots

# Training data

1

Formability classification

→

Training data: Experimental

- Training data compiled from literature and experimental databases ( ICSD etc)
- 1505 single and double oxide compositions
  - 1187 perovskites
  - 318 non-perovskites

2

Thermodynamic stability classification

→

Training data: DFT

- Criterion:
  - Energy above hull < 50 meV/atom
- 3271: stable
- 1881: unstable

3

Wide/narrow band gap classification

→

Training data: DFT

- Criterion: Band gap ( $E_g > 0.5$  eV)
  - 1575 : wide band gap
  - 3577: narrow band gap

4

Band gap regression

→

Training data: DFT

- 1575 : wide band gap materials

- Training data:
- Calculated using DFT
  - 5152 compounds
  - GGA - PBE formalism

# Features: Machine Learning models

## Combination of chemical and structural features

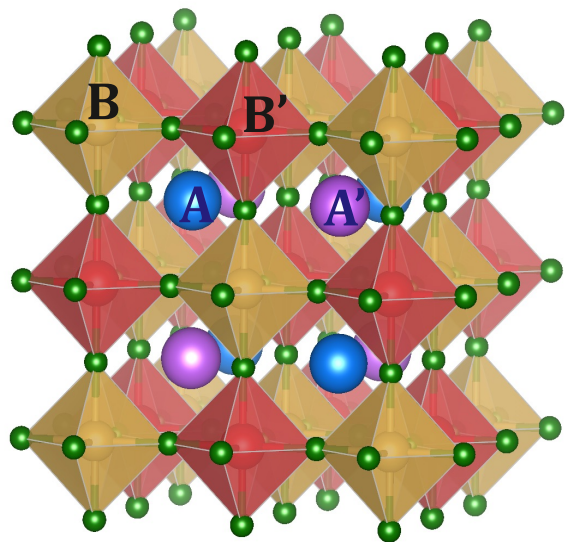
Pseudopotential radius*	Highest occupied atomic orbital ( HOMO)*	Tolerance factor ( $\tau$ )
Electronegativity*	Lowest occupied atomic Orbital ( LUMO)*	Octahedral factor ( $\bar{\mu}$ )
Electron affinity*	Ionization energy	Mismatch factor ( $\Delta\mu$ )

\* Element specific

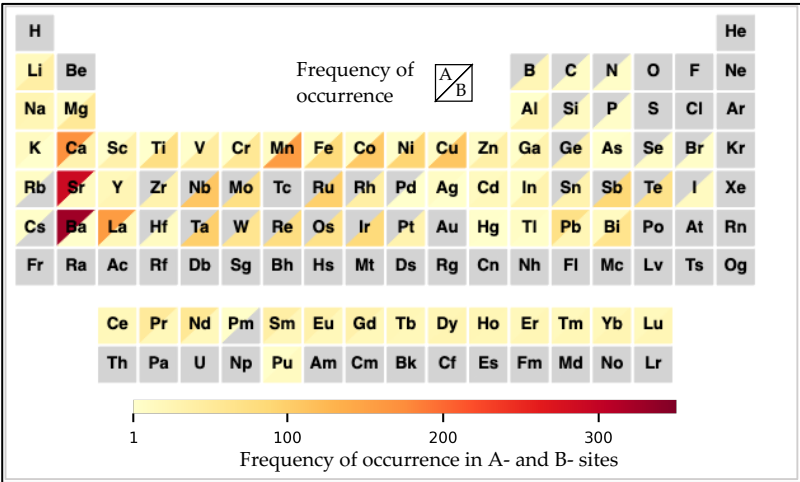
$$\tau = \frac{r_A+r_O}{\sqrt{2}(r_B+r_O)}$$

$$\bar{\mu} = \frac{r_B}{r_O}$$

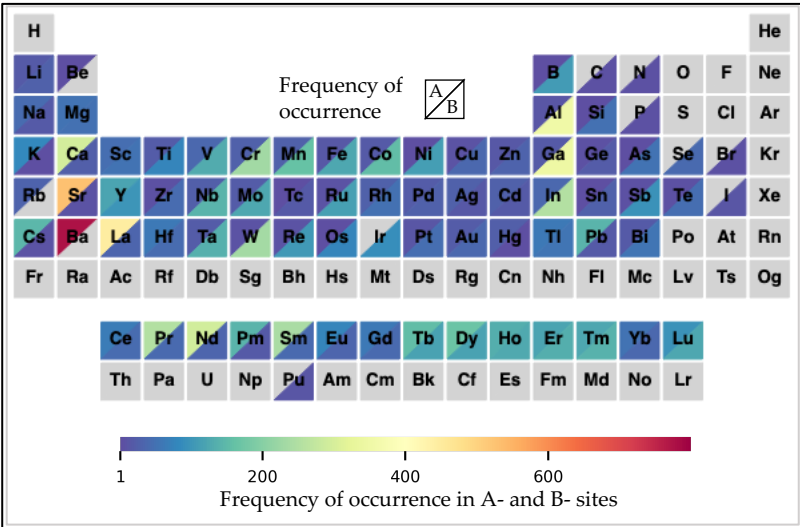
$$\Delta\mu = \frac{|r_{A/B} - r_{A'/B'}|}{r_O}$$



## Formability Training dataset



## DFT Training dataset



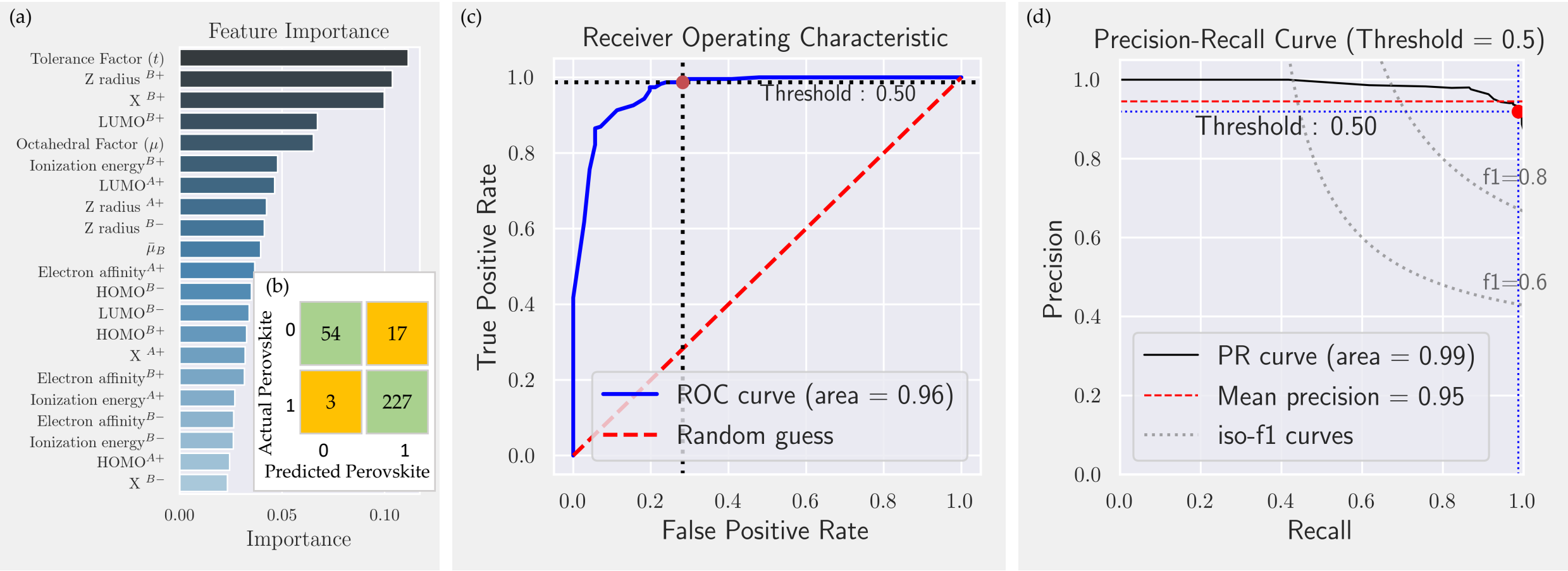
# Comparison of perovskite formability and stability

- Formability: *Ability to experimentally synthesize a model*
- Relies on geometric criteria derived using either ionic radii or bond distances and is a qualitative approach to identifying chemistries that will form perovskites.
- Stability: *Thermodynamic preference to form the structure*
- Energy hull construction to determine if the structure is on the convex hull and will stabilize and not decompose.

It is not known with certainty whether a formable perovskite is necessarily thermodynamically stable and vice versa.

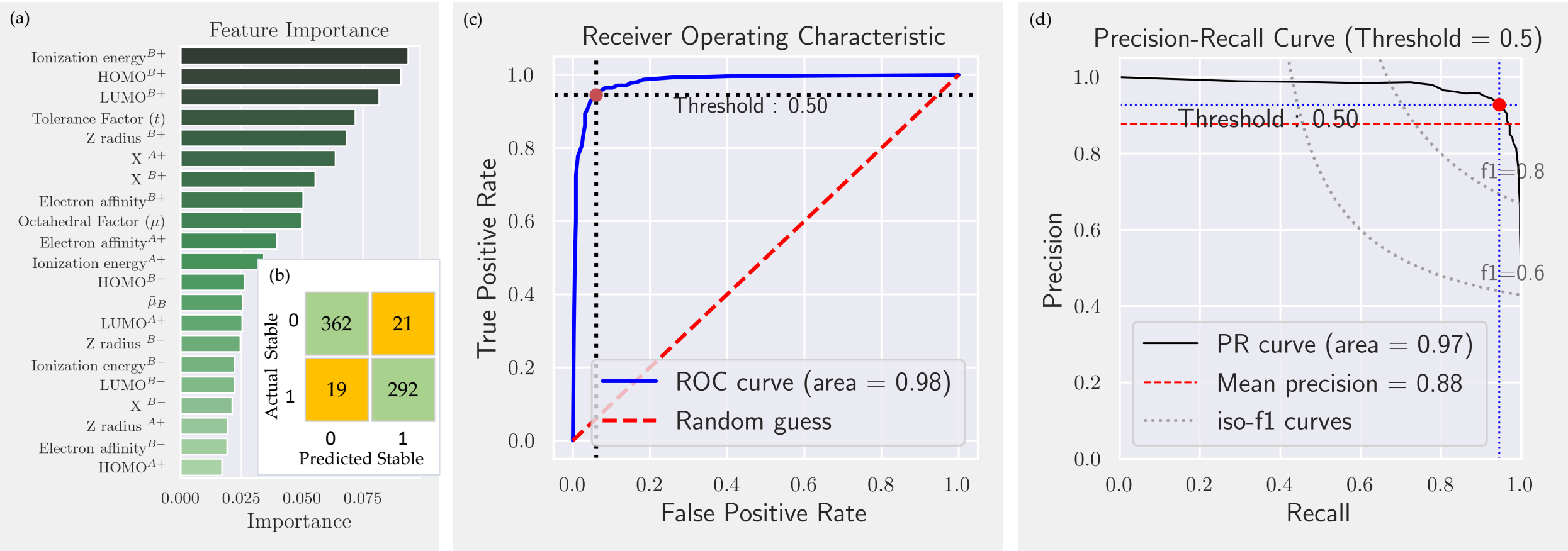
Are both formability and thermodynamic stability necessary to guarantee the viability of a composition as a perovskite candidate or is one a more robust metric compared to the other?

# Formability classification model



Random forest classification results for perovskite formability. a) Feature importance plot for all the features with non-zero values, b) Confusion matrix, c) Receiver operating characteristic (ROC) curves, and d) Precision-recall curves of the cross-validated random forest classification on test data.

# Stability classification model



Random forest classification results for perovskite stability. a) Feature importance plot for all the features with non-zero values, b) Confusion matrix, c) Receiver operating characteristic (ROC) curves, and d) Precision-recall curves of the cross-validated random forest classification on test data.



# Wide/narrow band gap classification model

<https://nanohub.org/tools/perovMLdis>

## ML-aided High-throughput screening for Novel Oxide Perovskite Discovery

By [Anjana Talapatra](#)  
*Los Alamos National Laboratory*

ML-based tool to discover novel oxide perovskites with wide band gaps

Edit

Launch Tool

Version 1.0 - published on 15 Jul 2021  
doi:10.21981/TWE2-ZE74 [cite this](#)  
This tool is closed source.  
[View All Supporting Documents](#)

3 users, detailed usage

0 Citation(s)

0 questions (Ask a question)

0 review(s)

0 wish(es) (New Wish)

Share: [f](#) [t](#) [in](#) ...

AboutUsageCitationsQuestionsReviewsWishlistVersionsSupporting Docs

Category

Tools

Published on

15 Jul 2021

Abstract

One of the most basic approaches to problem solving is to conceptualize the problem at different abstraction levels and translate from one abstraction level to the others easily, i.e., deal with them hierarchically. This concept is especially applicable to the field of novel materials discovery, wherein large candidate domains can be quickly and efficiently explored by hierarchically discarding irrelevant candidates. In this tutorial, we illustrate this approach using the example of wide band gap oxide perovskites. We will sequentially search a very large domain

Watch resource

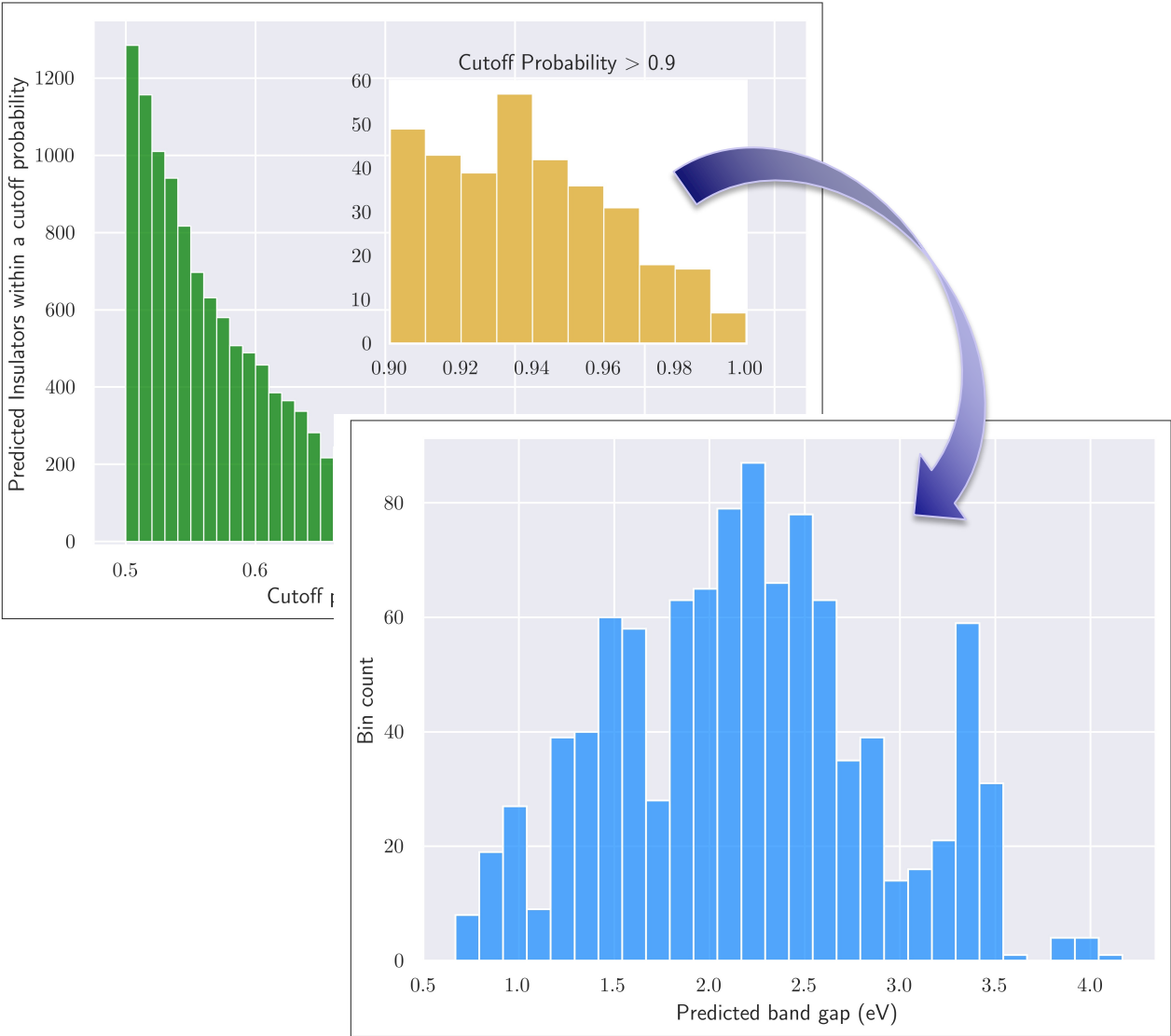
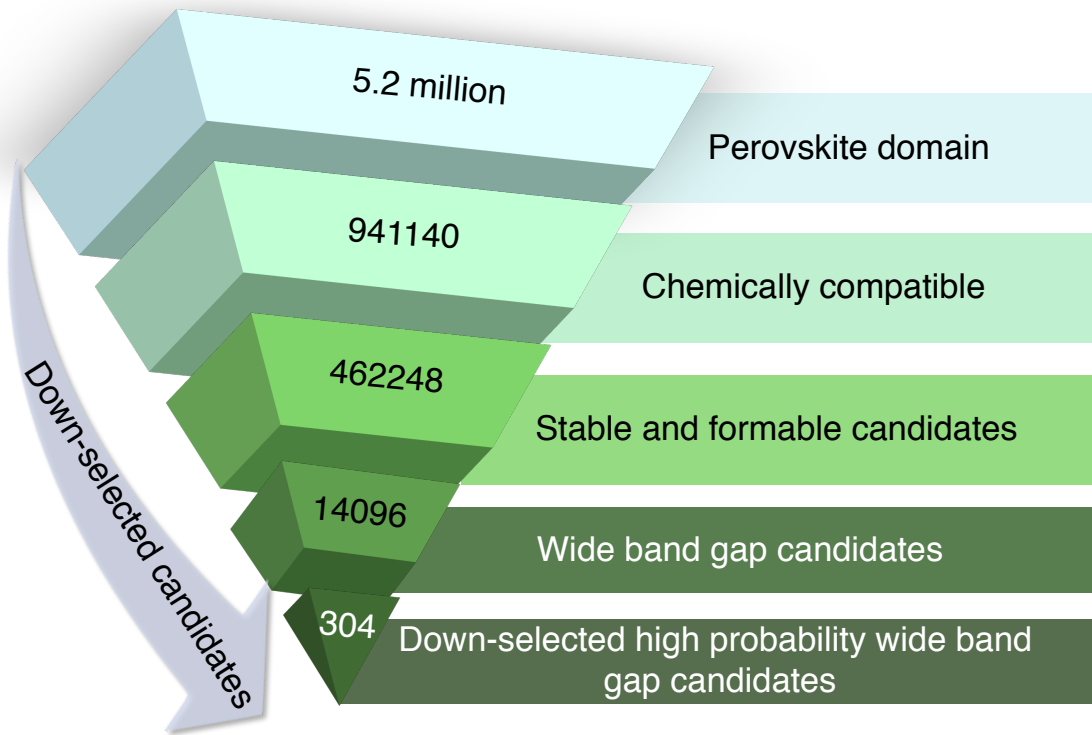
When watching a resource, you will be notified of updates. You may stop watching at any time.

UNCLASSIFIED

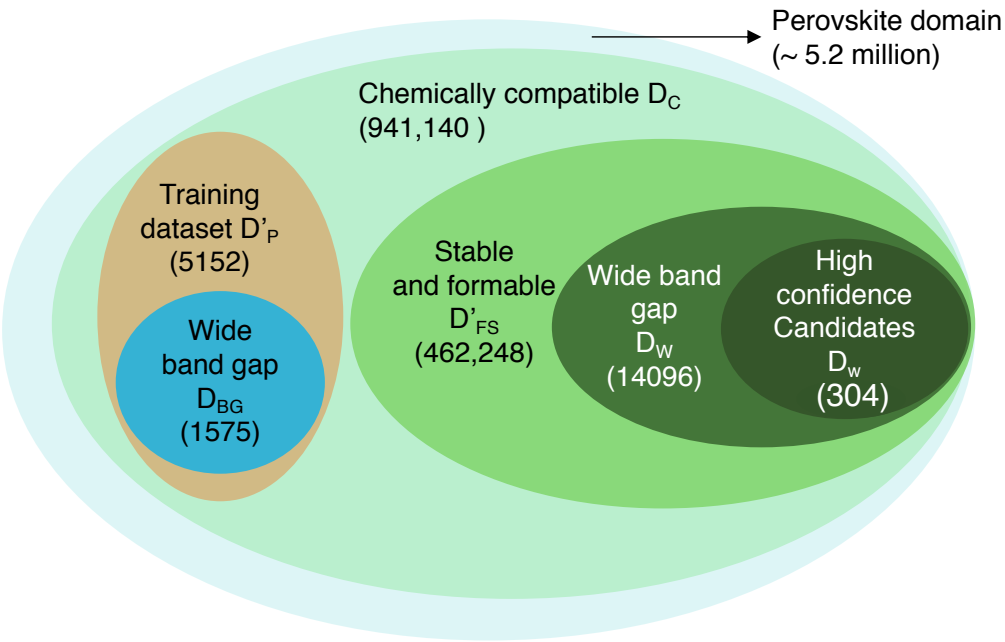
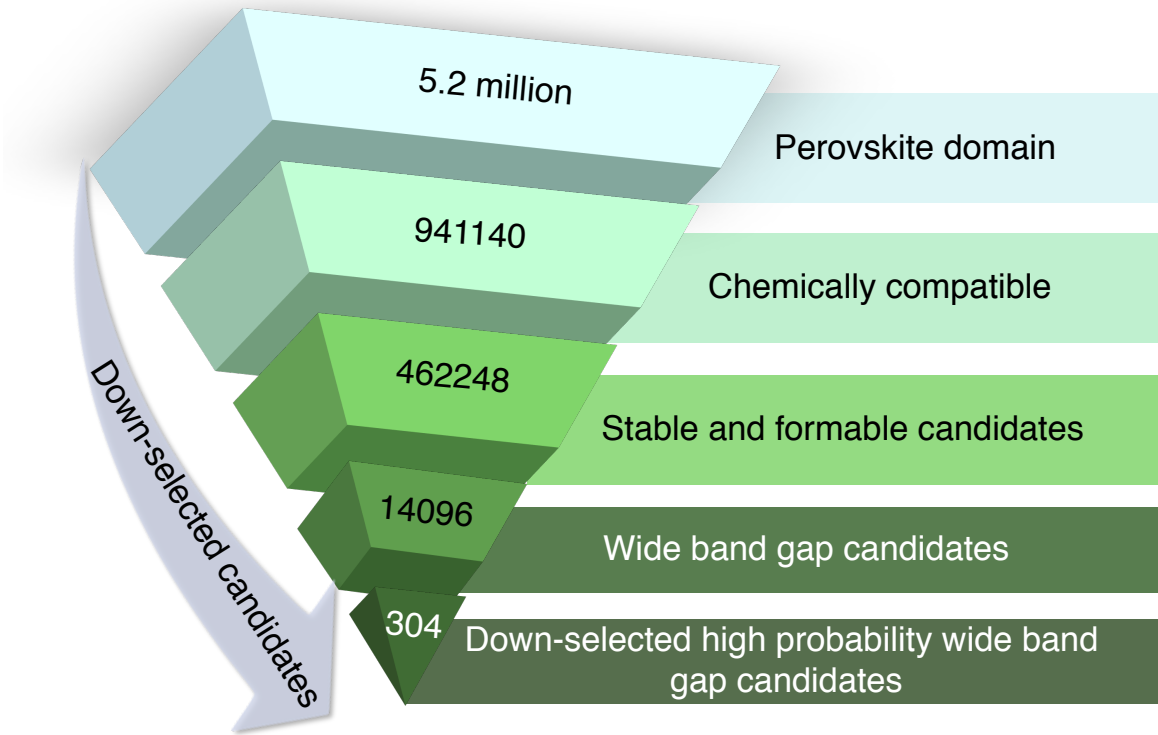
| 18



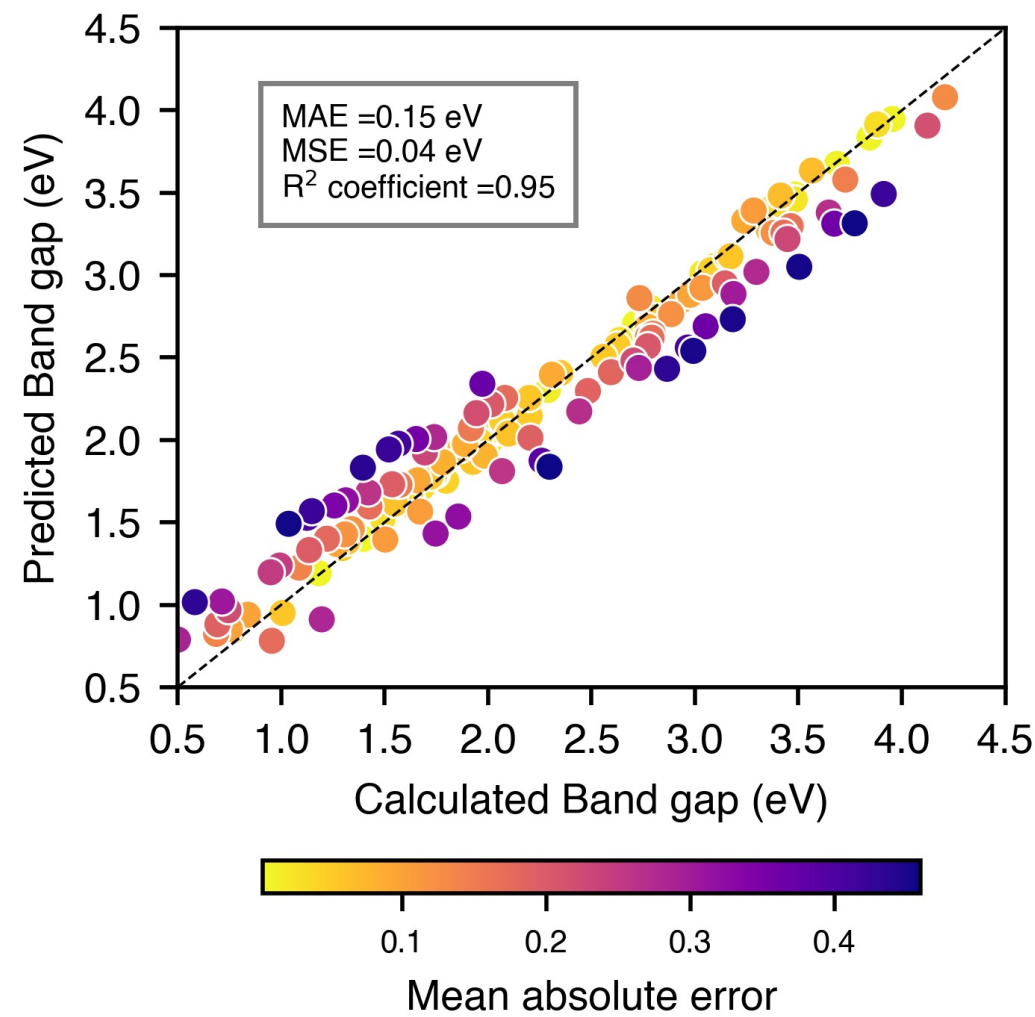
# Novel wide bandgap oxide perovskite predictions



# Novel wide bandgap oxide perovskite predictions



# Computational confirmation of results



- 150 of the predicted 304 candidates were randomly selected and DFT calculations carried out.
  - Wide band gap ✓
  - Calculated bandgaps accurate with a average MAE = 0.15 eV

## Some more suggestions

- Use machine learning only if necessary
- Ensure code is reproducible, with no ad-hoc measures, and all data sources annotated if applicable
- If permissible, have data and codes and scripts publicly available in a repository
- Answer emails from fellow researchers regarding published work and repositories

# Thank you!